NWS Operations Proving Ground

Operational Evaluation Report

Real Time Collaborative Forecast Process for QPF using Cloud AWIPS Experiment

An OPG Virtual Experiment

NWS Operations Proving Ground
7220 NW 101st Terrace
Kansas City, MO 64153

1. Executive Summary

In late January 2021, the Program Management Office (PMO) and the Operations Proving Ground (OPG) discussed the potential for hosting several (possibly weekly or bi-weekly) real time demonstrations of the Collaborative Forecast Process for QPF using AWIPS in the Cloud. The purpose of these demonstrations would be to provide forecasters a realistic experience of the CFP prior to the national demonstration planned for the fall (2021). The OPG felt this concept was within our capabilities, but we wanted to run a test with some field office volunteers to ensure we could meet the PMO's expectations.

Roughly a week later, the OPG ran a two hour test of the CFP with seven forecasters from around the CONUS and three OPG meteorologists. The OPG also invited representatives from the PMO, WPC, and OCLO to observe the test. We divided our ten participants into pairs who played the role of five WFOs in the southeast US (Tallahassee, Jacksonville, Atlanta, Birmingham, and Charleston, SC). The OPG selected the southeast US because there were several expected rain events over the seven day forecast period. None of the events were anticipated to be particularly significant, but some localized flooding or prolonged moderate rain was suggested by model guidance.

As usual, the participants performed exceptionally well. In this case, because of the short spin up time for the test, the participants did not receive any pre-test training (the OPG will often provide some training on tools or methods we expect participants to use during a test). They were provided with a document that described their role, expectations, instructions for accessing AWIPS in the Cloud, and chat best practices (further discussed below). Otherwise, the OPG simply said, "evaluate the environment as you normally would and inform your neighboring offices if you feel the NBM starting point needs any adjustments." From there, the participants had about two hours to complete their assessment, collaborate with neighboring offices, and attempt to agree on any changes to the NBM.

Somewhat surprisingly, the participants completed their task in about an hour and half. This allowed the OPG to engage in an unexpected short debrief session with the participants. Between the debriefing, and a short survey we asked participants to complete, the OPG learned a few important lessons we will apply to future CFP for QPF demonstrations. These lessons are discussed in short detail below. The most important lesson was that the OPG feels we can effectively conduct several more live CFP for QPF demonstrations using AWIPS in the Cloud for many forecasters across the NWS.

2. Experiment Design and Technical Specifications

As mentioned in the executive summary, ten volunteers joined our experiment (seven field office forecasters and three OPG staff members). The OPG spent the days leading up to the experiment monitoring weather conditions to select an area in the CONUS where a few interesting rain events were possible. We settled on the southeast US the day prior to the experiment. We grouped our participants into pairs and assigned each pair to one of the five CWAs in the southeast US noted in the executive summary. This allowed our participants the opportunity to review data prior to the test for their assigned CWA just as they would in real operations.

On the day of the test, we spent a few minutes on a Google Meet ensuring our participants had access to AWIPS in the Cloud and answered any last minute questions. Participants were instructed to conduct their atmospheric assessment as they normally would in operations, and then determine if the NBM QPF starting point required any modifications during the days 2 through 7 period. They were asked to collaborate on any desired changes to the NBM in Google Chat with their neighboring offices. We made it clear that a conference call/Google Meet was only necessary if the participants were unable to come to consensus via the chat collaboration. We also made it clear that the participants should only focus on QPF amounts - not flood headlines or any other forecast element.

One of our participants did experience minor technical problems resulting in a delay accessing Cloud AWIPS. However, this individual was eventually able to engage effectively in the evaluation. Otherwise, our participants all indicated that Cloud AWIPS functioned very well with no major issues.

3. Results

To properly contextualize the results of the evaluation, we want to provide the reader with a timeline of events:

- 10:15 am CST: Participants began analyzing data and collaborating with their WFO partner
- 10:51 am CST: First office-to-office collaboration message in Google Chat
- 11:30 am CST: Final Google Chat collaboration message

So, in roughly one hour and fifteen minutes, our participants assessed the NBM QPF, discussed desired changes, and came to consensus without requiring video or phone collaboration. About 40 minutes was devoted to reaching a consensus on desired changes to the NBM via chat collaboration.

3.1 Chat Collaboration Best Practices

Prior to the start of the exercise, participants were given a list “Google Chat Collaboration Best Practices” that was compiled by the OPG based on our previous CFP for QPF experiment in August of 2020. These “best practices” included points on keeping chats simple and to the point, not burying the lead, sharing images for clarity, and avoiding vague chat to name a few. See Appendix 1 for a complete list of Google Chat collaboration best practices provided by the OPG. The participants seem to effectively apply these chat room best practices which may have played a role in some of the noted differences in the Google Chat collaboration as compared to the August 2020 CFP for QPF experiment. .

For example, the excerpt from the Google Chat collaboration in Figure 1 exhibits several best practices. First, notice how Derrick initiated the conversation with some important context, followed by his meteorological reasoning, and finally an assessment of the NBM QPF (in this case being too low). This chat message, along with the others shared in Figure 1, exemplify the straightforward and clear communication that seems to achieve the most effective collaboration.

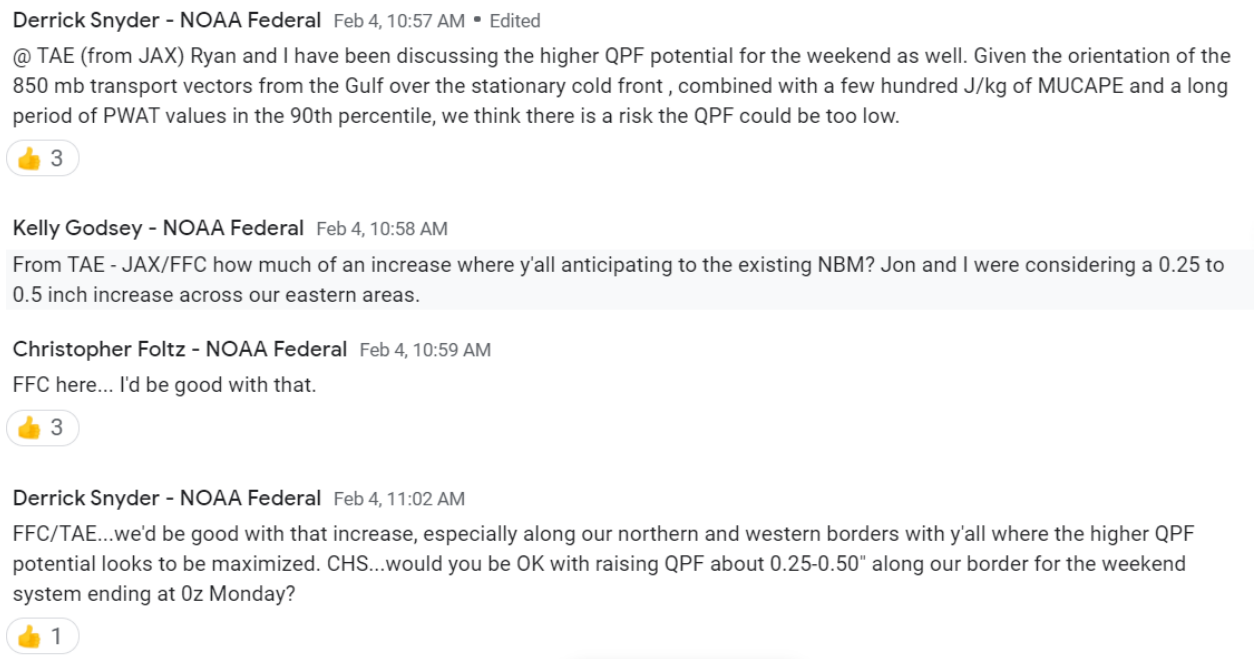

Figure 1: Screen capture of chat exchanges during the CFP for QPF experiment.

Also important are the three “thumbs up” notifications below Derrick’s chat (Figure 1). We believe it is reasonable to assume, given prior collaboration chat room assessments, that the three individuals who acknowledged their agreement with Derrick by giving a “thumbs up” would have normally written into chat something like, “Sounds good to us,” or “We agree.” The “Feedback” category, which would include statements such as “We agree,” had by far the most number of chats in the August 2020 CFP for QPF experiment. However, in the example from this test, Derrick was informed that three people agree with him without those three people taking up valuable chat space with additional chat messages. While reaction emojis were used in the August 2020 CFP for QPF experiment, we feel they may have been underutilized.

In the current test, participants also shared images that helped convey their reasoning for suggesting changes to the NBM starting point, as suggested in the best practices document. Interestingly, three of the four images presented in the Google Chat room were based on ensemble data (either raw or post-processed).

Finally, it is worth noting the low number of chats that occurred during the total exercise time frame, and in particular the lack of any chats for the first 35 minutes of the exercise. This finding is especially noteworthy given its contrast to the CFP for QPF experiment conducted in August 2020. In the August 2020 experiment, there were on average just over 16 chats per day (three days total) within the first 30 minutes that participants analyzed data. Of course, there were only five offices represented in the current test in contrast to seven offices in the August 2020 experiment, and three of those offices were not WFOs in the August 2020 experiment (WPC, NWC, and RFC). Thus, any direct comparisons should be made with caution. Nevertheless, it is possible that participants in the current experiment heeded the

advice of the best practices document which noted “if there is nothing to talk about, don’t clutter the chat.”

It is also believed that the experimental design of having two participants representing each office, and collaborating among themselves from the start, may have played a role in the decrease in chat messages. Whereas in the August 2020 experiment, only one of the seven offices had an office partner to collaborate with from the start. As such, given that in normal operations there are in-office colleagues to collaborate with, the large amount of chat in the August 2020 experiment, especially early in the collaboration period, may be somewhat unrepresentative of what would occur in an operational environment.

Finally, we are largely making an assumption that the collaboration experience was good in the current test based on the final outcome. The only data collected directly from participants on their Google Chat collaboration experience was when participants chose to note Google Chat in their response to a broad survey question on what went well. The survey responses that did comment on Google Chat noted its usefulness; two participants specifically commented on the benefit of sharing images. Nevertheless, without more specific feedback on the use of Google Chat, it is difficult to directly say that the chat experience was better in this test than in the August 2020 experiment. However, we can say that some of the negative, as noted by participant feedback, attributes of the August 2020 chat experience (i.e., chat volume, “running stream of consciousness”) were mitigated in this test.

3.2 On Workload and Collaboration Options

A key lingering question is, “how should WFOs convey their desired changes to WPC?” The current workflow for the CFP for QPF indicates offices will either exchange thoughts or use GFE to share their edit requests with WPC using the ISC feature. In this scenario, local forecasters would first analyze data, then collaborate with neighboring offices (and/or WPC) on proposed changes to the NBM, then modify their local GFE database according to the results of collaboration, and finally share their edited grids with WPC via ISC. The OPG views this as compartmentalizing the forecast process into distinct bins: analysis, collaboration, grid modification, grid sharing, WPC assessment, WPC modification, grid sharing, and final adjustments.

During our experiment, we noticed that this compartmentalization of the process would have resulted in increased workload in a real event. In this case, there were two offices in the experiment who felt the NBM QPF in the time frame of interest (days 2, 3, and 4) was representative, and thus, they were comfortable using the NBM QPF as the forecast. In other words, they would not have edited any grids. However, once the other three offices felt confident that the NBM needed to be adjusted up, the two offices who preferred to accept the NBM QPF unedited were now faced with making grid modifications. In reality, this could extend to several offices thanks to the grid editing “domino effect.”

Keep in mind, once those grids are edited and shared with WPC, WPC may in turn make additional modifications. If so, this would mean five individuals spent time editing grids at the WFO level, only to have a sixth individual (at WPC) make another modification. Of course, the OPG can not confirm this

hypothesis without running larger scale tests, but we believe there is sufficient anecdotal evidence to support the hypothesis.

Based on this experience, we feel there may be a more efficient methodology to streamline the first few steps of the process by de-compartmentalization. We believe we can combine analysis, collaboration, modification, and sharing into one step utilizing a combination of geospatial collaboration tools such as ArcGIS Online, and Google Chat/Meet for interaction as evidenced by feedback in the August 2020 CPF for QPF experiment.

If participants, including WPC, work collaboratively to highlight areas on a shared map where the NBM needs modifications, then WPC could make those initial adjustments and there would be no reason to modify local QPF grids as a collaboration method. Instead, local QPF grid modifications would focus on the short term during rapidly changing situations.

We plan to evaluate this methodology in future tests to help inform the CFP workflow or future AWIPS design elements.

3.3 Visualizing Collaboration

Our participants all seemed to agree in the exercise feedback that they needed a graphical way to convey the location and magnitude of their desired changes. In typical operations, this would be communicated via ISC grids in GFE. However, the OPG's Cloud AWIPS instance is currently only configured to display GFE for WFO Omaha. Thus, GFE was not able to be utilized during this test and, as a result, participants did not have an easy way to spatially convey their proposed edits to their neighboring WFOs. So it seems collaboration was less effective than it may have been if participants had a better way to collaborate spatially. Eventually one of our participants created a graphic with a map with their proposed edits to the NBM that helped convey what was being discussed in chat (Figure 2).

This issue was primarily localized only to this experiment given the OPG's GFE constraints. However, as discussed in section 3.2, the OPG feels that alternative methods of conveying desired changes to the NBM might be more efficient.

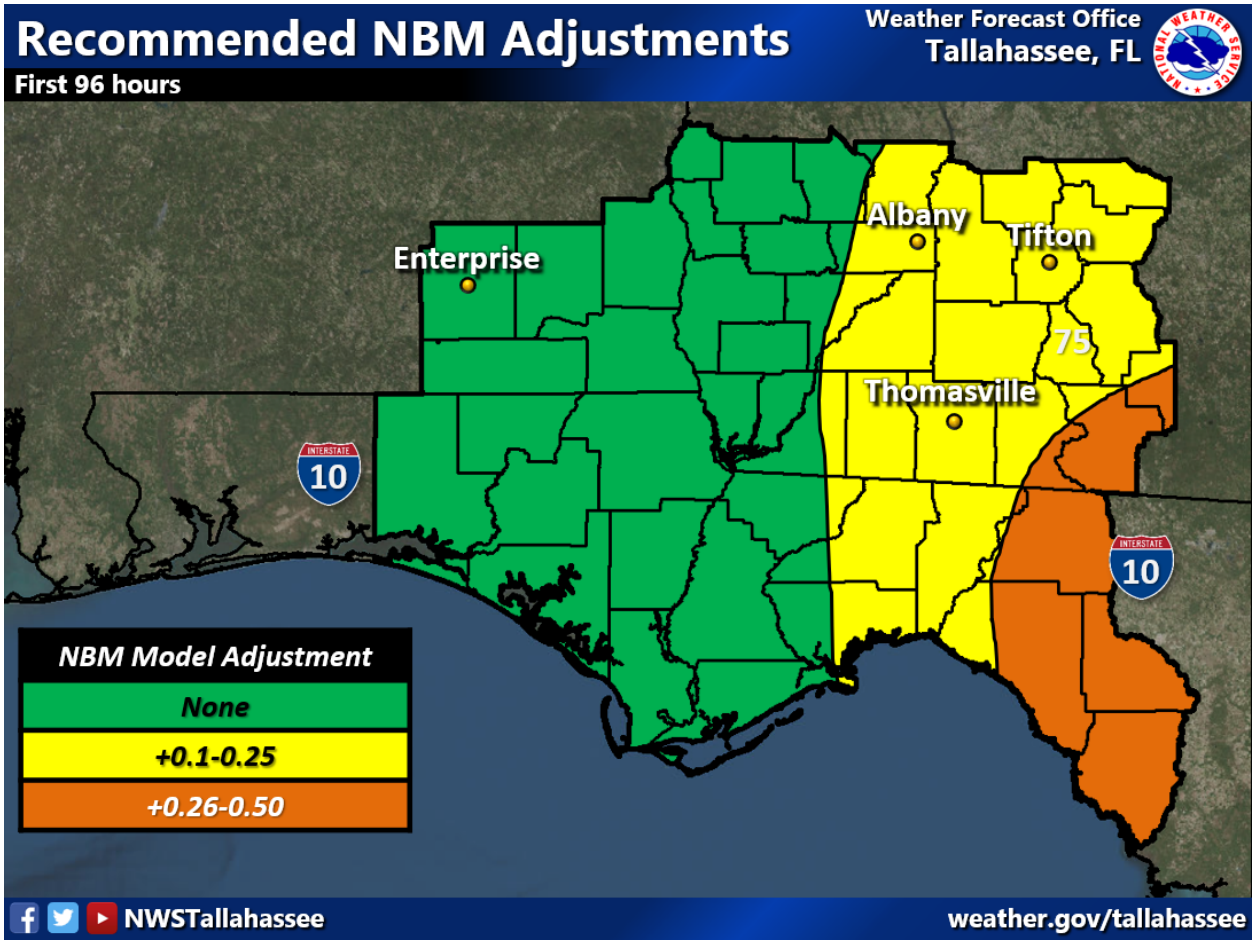

Figure 2: Graphic created by the simulated Tallahassee office highlighting the location and magnitude of desired changes to the NBM starting point.

Finally, one participant mentioned concerns about monitor space, "I only had one monitor, and struggled a little bit keeping track of a chat window, various tabs for guidance, the awips-in-the-cloud, the NBM viewer, and others." With future OPG experiments, we think participants will benefit from utilizing at least two monitors. This will allow the participant to run Cloud AWIPS on one screen, and utilize the second screen for collaboration (google chat / meets etc).

3.4 Use of AWIPS versus Web Content

Participants were given access to AWIPS in the Cloud for this test, which gave them access to most datasets in D2D that they would traditionally have in their offices during operations. However, essentially every participant noted that they utilized data found on the Internet, such as GEFS plumes, WAVE 1D Viewer, WeatherBell graphics, etc. more than data in Cloud AWIPS. While one participant mentioned being unable to use Cloud AWIPS after improperly logging out, and another mentioned not having their procedures and perspectives as a limiting factor, it seems that most participants were simply comfortable and satisfied with the data they had access to on the Internet.

Furthermore, some of the data that participants noted that they accessed on the Internet is not available in AWIPS (GEFS plumes, Ensemble Situational Awareness Tables, high resolution ECMWF ensembles). One participant asked, “was AWIPS in the cloud truly necessary?” That same participant added, “Our group (WFO) strictly used resources found on the Internet…” Another participant noted that the increase in supporting operations remotely during the COVID-19 pandemic has led more forecasters to find data on the Internet. Thus it is reasonable to assume that comfortability with accessing data on the Internet has increased as a result.

3.5 At What Point is QPF Worth Editing?

On a final note, forecasters decided the Day 3 to 4 NBM QPF needed to be increased by 0.25 to 0.5 inches in some areas. In other areas the changes were 0.1 to 0.25 inches. However, to some individuals these changes seemed unimportant (at least one participant noted that they preferred to leave the NBM alone instead of making a “small” change). In other words, some individuals might feel the NBM is “representative” despite potentially being “too low” or “too high” by half an inch. Participants discussed this topic in the debrief session. One forecaster felt it was important to make edits to the forecast in this scenario in order to stay on pace with what seemed to be an increasing trend in the QPF by the model data. In survey feedback one participant suggested potentially “developing thresholds for when changes should be done or when the NBM is good enough.” The question of what amount of QPF is worth editing seems to be an ongoing question; the answer certainly has implications to the CFP for QPF.

At the point, the OPG does not have sufficient evidence based on our experiments to suggest thresholds but we do recognize the need for guidance to the field in the near future.

4. Cost

As a reminder, in this case, ten individuals utilized Cloud AWIPS for approximately two hours. The participants joined our standalone AWIPS instance that provides 32 CPU cores, 244 GB of RAM and 2 graphics processors; each with 8 GB of RAM. They accessed the typical data sets common in forecast operations when diagnosing rainfall.

The average daily operating cost of our Cloud AWIPS instances during the month of February was about $43. On the day of our evaluation, our costs rose to $64 meaning this CFP for QPF evaluation cost the OPG about $21 above our typical expenses (roughly the cost of a fast food lunch for two people).

The OPG has now run several Cloud AWIPS tests, one of which had up to 24 continuous users of Cloud AWIPS. Granted, all of our tests have lasted less than two hours, but in all of our cases, the increase in our typical daily operating expenses have been nominal (on the order of $20 to $50).

5. Findings and Recommendations

With each small scale test, the OPG is growing in understanding and expertise in running real time experiments leveraging AWIPS in the Cloud. This is all made possible thanks to our virtual operating capacity, the dedication of our participants, and of course hard work by the OPG team.

Finding 1: The OPG believes we can conduct several additional live CFP for QPF demonstrations with ~20 field forecasters (each time) using Cloud AWIPS. We believe we can provide an effective real-time user experience to help forecasters become familiar with the CFP prior to the national demonstration.

Recommendation 1: The PMO and OPG should collaborate on a schedule, expectations, and roles/responsibilities to set up real-time CFP for QPF demonstrations.

Finding 2: The OPG believes that the use of two monitors improves the participant experience. While not required, two monitors allow users to display Cloud AWIPS on one monitor and web based collaborative tools (Google Chat/Meet) on another.

Recommendation 2: For future OPG evaluations/experiments, the OPG will work with the participants and help them acquire a second monitor from their local office (if needed).

Finding 3: The OPG believes that the "Collaboration Best Practices" information shared with participants helped improve the quality and efficiency of chat based collaboration. The OCLO is already working on training materials for forecasters to review prior to the national CFP demonstration.

Recommendation 3: Forecasters should review training materials provided by the OCLO regarding collaboration best practices. Forecasters should consider reviewing Crew Resource Management materials to further understanding of communication in high stress, time sensitive situations.

Finding 4: Participants noted a desire for geospatial collaboration tools during the experiment and specifically suggested the use of GFE as a form of geospatial collaboration. However, the OPG believes using GFE as a collaboration tool early in the collaborative forecast process may be inefficient and increase forecaster workload (although the OPG could not empirically prove this finding during one experiment).

Recommendation 4: The OPG, in collaboration with the PMO and WPC, should conduct follow-up CFP experiments focused on alternative methods to GFE for the exchange of thoughts early in the collaborative process. These methods should provide a geospatial collaboration component noted by our experiment participants.

Finding 5: Participants frequently used web based data analysis tools (such as WeatherBell or NWS supported web based tools) instead of utilizing data in AWIPS in their forecast process. The OPG acknowledges this finding is based on a very small sample size, but anecdotal evidence suggests that a growing number of forecasters are leveraging web based tools in their analysis process.

Recommendation 5: The OPG humbly suggests, if not already being done, for the NWS AWIPS program to reach out to forecasters who actively use web based tools for atmospheric analysis. The OPG could help facilitate these conversations if desired, but the end goal would be to

understand why forecasters prefer these web based tools and how we might incorporate their preferences into future AWIPS designs. Also, forecasters use of NWS supported web based tools, such as the WSUP viewer and NBM 1D viewer, lends reason to continue development of the tools.

Finding 6: The participants completed the first few steps of the Collaborative Forecast Process (analyze data, collaborate with neighbors, and share thoughts on necessary NBM changes) in about one and a half hours. This timeline matches extremely well with the approved CFP for QPF workflow.

Recommendation 6: The PMO, OPG, and WPC should investigate the possibility of conducting an end to end CFP for QPF test this summer (2021) using GFE within AWIPS in the Cloud.

6. Conclusion

OPG's evolving understanding of operating in a virtual capacity continues to improve. Our attempts at conducting shorter duration, focused, and real-time experiments have proven beneficial and productive. In this case, we showed our capacity to spin up a real time experiment in less than two weeks. We applied lessons learned from prior experiments by focusing on collaboration best practices. We identified potential inefficiencies with using GFE as a collaboration tool, and we learned that forecasters may prefer web based data analysis tools over AWIPS for rainfall forecasting. Of course, our sample size was far too low to comment on the ubiquity of forecasters' use of web based analysis tools.

The OPG is very excited to host additional CFP for QPF user experience sessions if the PMO feels these experiences are beneficial.

Appendix 1

Google Chat Collaboration Best Practices

1. Make chats short, simple, and to the point.

2. Don't bury important information (i.e.. Hey did you see the explosion from the S9 rocket? It was really cool to watch. I think the QPF for Day 3 is a bit off. Thoughts? How is your mother doing?)

3. Use video or phone conferencing to go into deep conversations, not chat. Otherwise, chats get too difficult to follow for other individuals.

4. Use upvoting/downvoting (thumbs up or down emoji) instead of messaging "I agree" or "I will look into this" to keep text short and sweet.

5. Focus on: Identifying discrepancies → analysis → decision making. Final decision making should be done via video or phone conferencing.

6. Use images and maps for clarity. They can be easily attached to a chat.

7. Avoid vague chats (i.e.. I think temps are a little low). Others might not know what area you are referring or what you think is "low."

8. Ask for clarity if a chat is vague. Don't assume things.

9. In a larger group setting, elect a "speaker" to spearhead things. They should help with focus and moving people through the collaboration process.

10. Just because a chat room exists doesn't mean you are **forced** to talk, if there is nothing to talk about, don't clutter the chat.

Google Chat Technical Tips

1. @mention individuals to get their attention and a response.@mention-ing is equivalent to placing someone in the To or CC field of an email. Use @ accordingly; don't overuse it.

2. Urgency in conversations is comparable to urgency of an email in normal situations. Don't expect an immediate response even if you @mention someone in a conversation. Expect the response time to vary just like it does with email.

3. Pressing the up arrow in the message box will load your last message for you to edit - great for fixing typos.

4. Shift-enter creates a new line, which is nice for longer posts.

5. Pressing “r” will reply to the Conversation you last selected.

6. Text Formatting:
    - Bold text by enclosing it in asterisks: *test*
    - Italicize text by enclosing it in underscores: _test_
    - Strikethrough text by enclosing it in tildes: ~test ~
    - Create a code block formatted (monospaced) text by enclosing it in ticks: `test`

You can combine some of these operators: _*test*_